

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-110385

(43)Date of publication of application : 23.04.1999

(51)Int.Cl.

G06F 17/27

(21)Application number : 09-268592

(71)Applicant : NIPPON HOSO KYOKAI <NHK>

(22)Date of filing : 01.10.1997

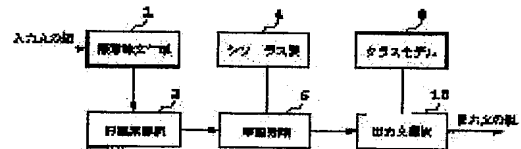
(72)Inventor : ANDO AKIO

## (54) DEVICE AND METHOD FOR PROCESSING LANGUAGE

(57)Abstract:

PROBLEM TO BE SOLVED: To deepen the degree of relation for a classified word as well and to improve the accuracy of chain probability by semantically classifying the word according to a thesaurus.

SOLUTION: A processing block 3 divides the pair of input sentences into respective three parts and by exchanging these parts, a non-semantic sentence is generated. A morpheme analytic block 2 divides the pair of generated non-semantic sentences for the unit of a word. A storage block 4 stores the thesaurus. A word classifying block 6 converts a word to a class code based on the thesaurus. A storage block 8 stores the table of trigram probability between classes. A processing block 10 calculates Perplexity based on the class trigram concerning the set of all the generated non-semantic sentences and selects the set of non-semantic sentences having the highest Perplexity. Thus, the classification of the thesaurus is stored and according to this classification of the thesaurus, the word is classified.



## \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

CLAIMS

---

[Claim(s)]

[Claim 1]A language processor with which one class classifies two or more words into two or more classes which comprise a set of two or more words, acquires chain probability which shows a grade of a chain between the classified classes concerned in them, and chooses a class and/or a word sequence as them based on the acquired chain probability concerned, comprising:

A memory measure which memorizes a classification of a thesaurus.

A sorting means which classifies said two or more words according to a classification of this thesaurus.

[Claim 2]A language processor when classifying said two or more words in the language processor according to claim 1, wherein said sorting means changes a word into a classification code and calculates said word sequence with a gestalt of a classification code sequence at the time of calculation of said chain probability.

[Claim 3]A language processor, wherein said sorting means makes this attached word itself a classification code about an attached word attached to said word in the language processor according to claim 2.

[Claim 4]A language processor, wherein said sorting means classifies every part of speech in the language processor according to claim 2 in the case of an independent word with which said word is not indicated to a classification of said thesaurus.

[Claim 5]A language processor having further a selecting means chosen as a meaningless sentence which uses a word sequence from which said chain probability serves as the minimum in an acoustic-sense experiment in the language processor according to claim 1.

[Claim 6]A language processor having further a selecting means which chooses a word sequence from which said chain probability serves as the maximum as an optimal candidate in voice recognition candidates in the language processor according to claim 1.

[Claim 7]In the language processor according to claim 1, it relates to calculation of chain probability between each word, A language processor characterized by presuming chain probability between said words based on appearing probability of said word, and chain probability between said classes when appearing probability of a word in a class and chain probability between classes are acquired and chain probability between words does not acquire directly.

[Claim 8]One class classifies two or more words into two or more classes which comprise a set of two or more words, In a language processing method which acquires chain probability which shows a grade of a chain between the classified classes concerned, and chooses a class and/or a word sequence based on the acquired chain probability concerned, A language processing method memorizing a classification of a thesaurus and classifying said two or more words according to a classification of this thesaurus.

[Claim 9]A language processing method changing a word into a classification code and calculating said word sequence with a gestalt of a classification code sequence in the language processing method according to claim 8 at the time of calculation of said chain probability when classifying said two or more words.

[Claim 10]A language processing method making this attached word itself into a classification code about an attached word attached to said word in the language processing method according to claim 9.

[Claim 11]A language processing method classifying every part of speech in the language processing method according to claim 9 in the case of an independent word with which said word is not indicated to a classification of said thesaurus.

[Claim 12]A language processing method, wherein said chain probability chooses a word sequence used as the minimum in the language processing method according to claim 8 as a meaningless sentence used in an acoustic-sense experiment.

[Claim 13]A language processing method, wherein said chain probability chooses a word sequence used as the maximum as an optimal candidate in voice recognition candidates in the language processing method according to claim 8.

[Claim 14]In the language processing method according to claim 8, it relates to calculation of chain probability between each word, A language processing method characterized by presuming chain probability between said words based on appearing probability of said word, and chain probability between said classes when appearing probability of a word in a class and chain probability between classes are acquired and chain probability between words does not acquire directly.

---

[Translation done.]

## \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

## DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention]In natural language processing, this invention relates to the language processor and method of choosing / generating the optimal word sequence, and relates to the language processor and method using statistical information between the word classes especially set up based on the thesaurus.

[0002]

[Description of the Prior Art]Conventionally, the word n-gram model using the statistical chained information of n words was used in the language processor. The word was classified for every part of speech, the word class was created, and the class n-gram model using the statistics chained information between these word classes (between parts of speech) was also used (about the above, it is referring to north other "spoken language processing" Morikita Shuppan, for example). In these methods, when the number of words to treat increased for example, in the case of the word n-gram model and n became large, the problem called what is called a sparse problem it becomes impossible to fully presume the chain probability between words arose. On the other hand, in the case of the class n-gram model between parts of speech, there was a fault that only rough language processing could be performed.

[0003]

[Problem(s) to be Solved by the Invention]In an acoustic-sense mental experiment etc., although the experiment which judges whether a test subject can catch a sound exactly is conducted, there is a demand of liking to remove the influence of a semantic analogy and to investigate only the original auditory characteristic, in this case. There is a way the right creates grammatically the sentence along which a meaning does not pass as the method of satisfying such a purpose (for example, Kazuo Watabe: "effect to the advanced age hearing impairment person of the speech rate delay by rate of speech conversion", an Oto-Rhino-Laryngological Society of Japan, Inc. report, Vol.99, pp.445-453 (1996)). However, it was unable not to find the good method of creating the examination sentence along which a meaning does not pass, but to remove a maker's arbitrary element from the created examination sentence as a matter of fact.

[0004]On the other hand, in speech recognition, in order to choose from the recognition result in a sound level linguistically the word sequence which will seemingly be reasonable in order to raise recognition performance, or to reduce the number of the candidates who can assume as a recognition result after it using the recognition result obtained till then, the word n-gram is used. In the large lexical continuous speech recognition, especially use of the word n-gram is made indispensable. Also in natural language processing other than speech recognition, in order to choose the word sequence which will seemingly be reasonable out of two or more word sequences, the word n-gram is used. However, in a word n-gram model, when the case of n= 3 (the chain model of three words: trigram model) is taken for an example, for example and the number of words is 20,000 words. It is necessary to consider the combination of the cube (8,000,000,000,000) individual of 20,000, and a huge memory is needed in the case of processing, and also there is a problem that that statistics value cannot be presumed, from an available language database (this problem is called the spar SUNESU problem). As for this problem, n becomes much more remarkable, when larger than 3. Such a problem had had big influence on the processing result in speech recognition.

[0005]Then, the purpose of this invention is to provide the possible language processor and method of raising the accuracy of the chain probability between words, and easing the restriction about the memory of a language processor rather than classifying a word for every part of speech.

[0006]

[Means for Solving the Problem]In order to attain such a purpose, this invention an invention of claim 1, One class classifies two or more words into two or more classes which comprise a set of two or more words, chain probability which shows a grade of a chain between the classified classes concerned is acquired, and a language processor which chooses a class and/or a word sequence based on the acquired chain probability concerned is characterized by comprising the following:

A memory measure which memorizes a classification of a thesaurus.

A sorting means which classifies said two or more words according to a classification of this thesaurus.

[0007]In the language processor according to claim 1, said sorting means changes a word into a classification code, and an invention of claim 2 calculates said word sequence with a gestalt of a classification code sequence at the time of calculation of said chain probability, when classifying said two or more words.

[0008]An invention of claim 3 makes this attached word itself a classification code in the language processor according to claim 2 about an attached word to which said sorting means is attached to said word.

[0009]As for an invention of claim 4, in the case of an independent word for which said word is not indicated to a classification of said thesaurus, in the language processor according to claim 2, said sorting means classifies every part of speech.

[0010]An invention of claim 5 was further provided with a selecting means chosen as a meaningless sentence which uses a word sequence from which said chain probability serves as the minimum in an acoustic-sense experiment in the language processor according to claim 1.

[0011]An invention of claim 6 was further provided with a selecting means which chooses a word sequence from which said chain probability serves as the maximum as an optimal candidate in voice recognition candidates in the language processor according to claim 1.

[0012]An invention of claim 7 relates to calculation of chain probability between each word in the language processor according to claim 1, When appearing probability of a word in a class and chain probability between classes are acquired and chain probability between words does not acquire directly, chain probability between said words is presumed based on appearing probability of said word, and chain probability between said classes.

[0013]One class an invention of claim 8 in two or more classes which comprise a set of two or more words. In a language processing method which classifies two or more words, acquires chain probability which shows a grade of a chain between the classified classes concerned, and chooses a class and/or a word sequence based on the acquired chain probability concerned, A classification of a thesaurus is memorized and said two or more words are classified according to a classification of this thesaurus.

[0014]In the language processing method according to claim 8, when classifying said two or more words, an invention of claim 9 changes a word into a classification code, and calculates said word sequence with a gestalt of a classification code sequence at the time of calculation of said chain probability.

[0015]An invention of claim 10 makes this attached word itself a classification code about an attached word attached to said word in the language processing method according to claim 9.

[0016]In the case of an independent word with which said word is not indicated to a classification of said thesaurus, an invention of claim 11 classifies every part of speech in the language processing method according to claim 9.

[0017]In the language processing method according to claim 8, said chain probability chooses an invention of claim 12 as a meaningless sentence which uses a word sequence used as the minimum in an acoustic-sense experiment.

[0018]An invention of claim 13 chooses a word sequence from which said chain probability serves as the maximum as an optimal candidate in voice recognition candidates in the language processing method according to claim 8.

[0019]An invention of claim 14 relates to calculation of chain probability between each word in the language processing method according to claim 8, When appearing probability of a word in a class and chain probability between classes are acquired and chain probability between words does not acquire directly, chain probability between said words is presumed based on appearing probability of said word, and chain probability between said classes.

[0020]

[Embodiment of the Invention]Hereafter, with reference to drawings, the embodiment of this invention is described in detail.

[0021]First, the language processing method which applied this invention is described. According to this embodiment, the class n-gram model based on a thesaurus is used. A word n-gram model is a model which approximates occurrence of a word sequence in n-1-fold Markov course. Therefore, word sequence  $w_1 w_2 \dots$  It is a formula about the probability that word  $w_m$  will occur when  $w_{m-1}$  is given. [0022]

[Equation 1]

$$P(w_m | w_1 w_2 \dots w_{m-1})$$

$$\approx P(w_m | w_{m-n+1} w_{m-n+2} \dots w_{m-1})$$

[0023]It comes out and approximates (however,  $m \geq n$ ). Now, each word shall belong only to one class, respectively. The class to which word  $w_i$  belongs is expressed with  $c_i$ . At this time, it is word sequence  $w_1 w_2$  with a class n-gram model. -- The probability that word  $w_m$  will occur under the conditions that  $w_{m-1}$  was given is approximated with a following formula. : [0024]

[Equation 2]

$$P(w_m | w_1 w_2 \dots w_{m-1})$$

$$\approx (P(c_m | c_{m-n+1} c_{m-n+2} \dots c_{m-1}) \cdot P(w_m | c_m))$$

[0025]Formula which simplified several 2 formulas in the examination sentence generation for an acoustic-sense mental experiment : [0026]


[Equation 3] $P(c_m | c_{m-n+1} c_{m-n+2} \dots c_{m-1})$

It \*\*\*\*\*.

[0027]A word is classified into a class according to a class n-gram model, the probability (chain probability) which shows the grade of the chain between classes, and the appearing probability of the word in a class are used, and it is possible to reduce a number of parameters substantially compared with a word n-gram model. For example, when the number of V and classes is set to C for the size (the number of words) of the vocabulary to treat, the number of the independent parameters in the word n-gram is  $V^{n-1}$ . On the other hand, the number of the independent parameters of a class n-gram model serves as  $C^{n-1} + V - C$ . The number of the independent parameters at the time of being referred to as  $V = 20,000$  and  $C = 1,000$  is shown in Table 1.

[0028]

[Table 1]

 ID=000005
---

[0029]Although the method of setting up a word class as a class n-gram model based on a part of speech from the former, etc. were proposed, there was a problem that only a rough word classification could be performed. Although the method of classifying a word using the mutual information between words is also proposed, the problem of the algorithm which searches for the optimal classification not existing is known. The thesaurus is known as a technique for on the other hand performing the semantic classification of a word. So, in an embodiment, solution of problems in a word n-gram model, such as a spar SUNESU problem, is aimed at by using the class n-gram model based on this thesaurus.

[0030]For this reason, the example which made the meaningless sentence of listening comprehension checking for \*\* acoustic-sense mental experiment generate (a 1st embodiment)

\*\* The example of management to the spar SUNESU problem in the word n-gram model for speech recognition (a 2nd embodiment)

The case of two \*\* is illustrated. The case (trigram (chain probability of a triad)) where n of n-gram is set to 3 in any case is shown.

[0031]1st and 2nd embodiments show the example using the classification lexical table of the National Japanese Language Research Institute as a thesaurus. In the classification lexical table, five digits divided by the period, such as 1.3062, are given as a code number of each word. Since the number of classes will become large too much if this code number is made into the number which expresses one class as it is, after omitting the last single figure, by this embodiment, a set of the word which has the same code number is treated as one class.

[0032]In a classification lexical table, since only the independent word is registered fundamentally, when using the class model of several 2 formulas, the handling of an attached word poses a problem. How the word which is not registered into a classification lexical table is treated poses a problem. About the attached word, although it was also possible to have constituted a class by classifying grammatically, by 1st and 2nd embodiments, the method of making one class constituting from one word was adopted. About the word which is not registered into a classification lexical table, the method of registering automatically based on a morphological-analysis result was adopted at the time of class model learning.

[0033]Study of the word n-gram usually presumes the probability value of several 1 formula from the frequency of a word chain, after dividing the sentence in a language database per word by a morphological analysis. Class n-gram and study are performed in the same procedure.

[0034]First, a word is changed into a class code by dividing a language database per word and comparing with a classification lexical table. In that case,

an attached word does not perform conversion in a code, but makes the word itself a class code. The following processings are performed about the independent word which is not indicated in a classification lexical table. First, about conjugated forms, such as a verb, after comparing by a stem of a word, the conjugated form and a corresponding classification code number are added to the word-class conversion table prepared independently. A numeral also adds the numeral to a word-class conversion table, after assigning the code of a numeral compulsorily. About a noun, according to a morphological-analysis result, it distributes to the class of a "common noun", a "SA strange noun", a "proper noun", a "formal noun", an "adverbial noun", and a "tense noun", it doubles with it, and the addition to a word-class conversion table (table showing the correspondence relation of the word contained in a class and its class) is performed. It classifies into a strange class about the other word.

[0035]The news manuscript database of the broadcasting station was used for study of the word trigram and the class trigram. This puts in a database the news manuscript which the journalist of the broadcasting station drew up from 1992 to May, 1996. In study of the class trigram, the size of the vocabulary was made into 20,000 words and the other word was assigned to the class only for an unknown word. The class only for this unknown word differs from a strange class (although the former is a class of the word which is not contained in a vocabulary and the latter is a word contained in the range of a vocabulary, the class which belongs is a set of a strange word). The number of the obtained classes is 936 and 518 of them classes were created from the classification lexical table. The number of average words assigned to each class to an independent word was 41. On the other hand, it creates on conditions with the same said of the word trigram.

[0036](A 1st embodiment) Example \*\* is explained in detail hereafter. When doing an audio listening comprehension examination, in order to investigate how [ that was caught certainly ] it is, the incomprehensible sentence which consists of a plain word is created, and it is possible to experiment using the sound which read out this sentence. Hereafter, it illustrates about generation of such a meaningless sentence.

[0037]The short owner meaning sentence prepared beforehand is divided into three portions, respectively, and creation of a meaningless sentence is performed by replacing these portions. The example of an owner meaning sentence is shown below.

[0038]By a motorcycle A cigarette. It does not set. From this morning A gear tooth. It is painful. In a drawer It is locked. On the head A pigeon. It stopped. It is high. A building. It is in large numbers.

[0039]However, after changing each sentence into a class code sequence using a classification lexical table and a word-class conversion table for facilitation, a meaningless sentence is created, when the occurrence probability for every sentence is calculated and probability chooses most the group of the sentence which became low only using the trigram probability between the classes of several 3 formulas. In order for comparison to show the effect of this invention, same processing was performed using the word trigram. Below, a result is displayed instead of a probability value for convenience using Perplexity. Perplexit(ies) in this case are  $P(w_1 \cdots w_n)$  and its entropy about the occurrence probability of each sentence (word sequence  $w_1 \cdots w_n$  it expresses with  $w_n$ ). [0040]

[Equation 4]

$$H(L) = - \sum_{w_1 \cdots w_n} \frac{1}{n} P(w_1 \cdots w_n) \log P(w_1 \cdots w_n)$$

[0041]When it carries out, [0042]

[Equation 5]It is the quantity expressed with  $PP=2^{H(L)}$ . Therefore, Perplexity becomes large, so that the occurrence probability of a sentence is small.

[0043]The functional constitution of the language processor which performs language processing for the meaningless sentence generation of listening comprehension checking with such a language processing method is shown. In drawing 1, 1 is a processing block which generates a meaningless sentence by dividing into three portions the group (5 of the owner meaning sentences shown previously) of an input sentence, respectively, and replacing these. In this case, the number of the sentences generated is 14,400.

[0044]2 is a morphological-analysis block which divides the group of the generated meaningless sentence per word. 4 is a storing block which stores a thesaurus. 6 is a word classification block which changes a word into a class code based on a thesaurus.

[0045]8 is a storing block which memorizes a table of trigram probability (probability value of several 3 formulas) between classes. 10 is a processing block which calculates Perplexity based on a class trigram about a group of all the generated meaningless sentences and as which Perplexity chooses most a group of a meaningless sentence which became large.

[0046]Drawing 2 shows the contents of language processing performed in a language processor. Actual language processing is explained according to drawing 2. The contents of execution in each block of drawing 2 are as follows.

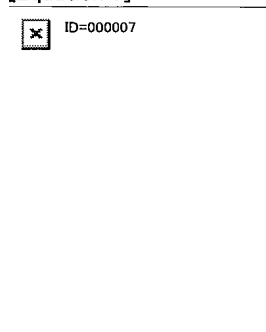
[0047]B-2: A block which reads a group of an input sentence.

[0048]B4: A block which generates a meaningless sentence by dividing each input sentence into three portions, fixing the 1st portion, and replacing the 2nd and 3rd portion. Five sentences, [0049]

[Equation 6]

. Express with  $b_{11}b_{12}b_{13}b_{21}b_{22}b_{23}b_{31}b_{32}b_{33}b_{41}b_{42}b_{43}b_{51}b_{52}b_{53}$ . ( $b_{ij}$  is the  $j$ -th block of the  $i$ -th sentence). The function  $r$  shall be fulfilled for  $i=j \Rightarrow r(i) = r(j)$  by the function from a set  $\{1, 2, 3, 4, 5\}$  to  $\{1, 2, 3, 4, 5\}$ . The function  $s$  is defined similarly. At this time, it is by this block, [0050]

[Equation 7]



[0051]It comes out and the sentence expressed is generated.  $r_i$  and  $s_j$  are the brief sketches of  $r(i)$  and  $s(i)$  here, respectively.

[0052]B6: The block which substitutes zero for the variable Max.

[0053]B8: The block which judges whether the group of all the meaningless sentences generated by B4 was chosen.

[0054]B10: The block which chooses the group which has not been chosen yet from the groups of the meaningless sentence generated by B4.

[0055]B12: A subroutine which calculates the Perplexity to a group of a meaningless sentence selected by B10. For details, it is shown in drawing 3. An input is a group of a meaningless sentence and a calculation result is returned to the variable out.

[0056]B14: A block with which a value of the variable out judges that it is larger than the variable Max.

[0057]B16: A block with which a value of the variable out judges that it is equal to the variable Max.

[0058]B18: A block which stores a group of a meaningless sentence which initializes an array which stores a group of an output statement and is

chosen there now.

[0059]B20: A block which replaces a value of the variable Max with a value of the variable out.

[0060]B22: A block which adds a group of a meaningless sentence chosen as arrangement which stores a group of an output statement now.

[0061]B24: A block which outputs a group of an output statement stored.

[0062]Drawing 3 shows the contents of processing in a subroutine shown in B12 of drawing 2. A processing block of drawing 3 is constituted by the following blocks.

[0063]B32: Block who substitutes zero for the variable out.

[0064]B34: Block who judges that all the sentences were chosen when choosing a sentence from a group of an inputted meaningless sentence.

[0065]B36: Block who chooses a sentence which has not been chosen yet from groups of an inputted sentence.

[0066]B37: Morphological-analysis Block who divides a selected sentence per word. As a morphological-analysis method, although what kind of method could be used, by this embodiment, the software Juman developed in Kyoto University was adopted.

[0067]B38: Block who decomposes a sentence into 3 groups of a word. When a sentence is made of connection of word  $w_1, w_2, \dots, w_n$ , 3 group  $w_1w_2w_3$  and  $w_2w_3w_4, \dots, w_{n-2}w_{n-1}w_n$  are generated.

[0068]B40: Block who classifies each word in a three word group into a class, and changes into three groups of a class code. Group  $w_1w_2w_3$  of three words,  $w_2w_3w_4, \dots, w_{n-2}w_{n-1}w_n$ . It changes into three class group  $c_1c_2c_3$  and  $c_2c_3c_4, \dots, c_{n-2}c_{n-1}c_n$ .

[0069]B41: A block which reads trigram probability of several 3 formulas about three each group generated by B40.

[0070]B42: From trigram probability read by B40 to a formula [0071]

[Equation 8]



ID=000008

[0072]The block which calculates Perplexity of a sentence as be alike.

[0073]B44: The block which substitutes for the variable out the sum of the value of the variable out, and the value of Perplexity calculated by B42.

[0074]B46: The block which outputs the value of the variable out.

[0075](A 2nd embodiment) Example \*\* is explained in detail below. Since 3 groups of the word contained in the uttered sentence generally did not exist at the time of study when using a word trigram model by speech recognition, the probability value may not be able to be presumed. In example \*\*, the probability value is presumed to such a case using the class trigram of several 2 formulas.

[0076]Drawing 4 shows the functional constitution of the language processor which performs language processing of example \*\*. In this example, there is a certain connected speech recognition system, and the case where this system outputs two or more recognition candidate sentences is assumed. In drawing 4, the morphological-analysis block which divides two or more sentences into which 22 was inputted as voice recognition candidates per word, and 24 are storing blocks which memorize word trigram probability (several 1 formula). 26 is a processing block which reads the word chain probability corresponding to an input. The storing block for which 28 stores a thesaurus, and 30 are word classification blocks which change a word into a class code based on a thesaurus.

[0077]32 is a storing block which memorizes a table of trigram probability between classes, and class word probability (probability that a word will appear in the class). 34 is a processing block which calculates class trigram probability (probability of several 2 formulas) corresponding to an input sentence. 36 is a processing block which calculates Perplexity to an inputted sentence (recognition candidate) and to which the value chooses and outputs the smallest thing (what has a high probability value), using class trigram probability, when word trigram probability is not called for.

[0078]Drawing 5 shows the contents of language processing in Example 2. Actual language processing is explained according to drawing 5. The contents of execution in each block of drawing 5 are as follows.

[0079]B-52: A block which reads a group (two or more recognition candidates) of an input sentence.

[0080]B54: A block which assigns the smallest possible value to a value of the variable Min.

[0081]B56: A block which judges whether all inputted sentences were chosen.

[0082]B58: A block which chooses a sentence which is not chosen by then.

[0083]B60: A subroutine which calculates Perplexity to a selected sentence. An input is the selected sentence and returns Perplexity to it using the variable out.

[0084]B62: A block with which the variable out judges whether it is smaller than the variable Min.

[0085]B64: Initialize an array which stores an output statement and store a sentence chosen as the arrangement now.

[0086]B66: Assign a value of the variable out to the variable Min.

[0087]B68: Output a sentence stored in arrangement for output statements.

[0088]Drawing 6 shows the contents of processing of the subroutine B60. A processing block of drawing 6 comprises the following blocks.

[0089]B72: A block which substitutes zero for the variable out.

[0090]B73: A morphological-analysis block which divides an inputted sentence per word.

[0091]B74: A block which decomposes a sentence into 3 groups of a word. Sentences are word  $w_1$  and  $w_2, \dots$  When made of connection of  $w_n$ , 3 group  $w_1w_2w_3$  of a word,  $w_2w_3w_4, \dots, w_{n-2}w_{n-1}w_n$  are generated.

[0092]B76: A block which judges whether 3 groups of all word were chosen.

[0093]B78: A block which chooses a three word group which is not chosen by then.

[0094]B80: A block which judges whether a word trigram exists to selected 3 groups.

[0095]B82: A block which adds an opposite numerical value of probability called for from a word trigram model to the variable out.

[0096]B84: A word classification block which asks for 3 groups of a class from 3 groups of a word chosen.

[0097]B86: When making into  $c_i$  a word class to which  $w_{m-2}w_{m-1}w_m$  and word  $w_i$  belong 3 groups of a word, Chain probability over 3 group  $c_{m-2}c_{m-1}c_m$  of a class, A block which calculates an opposite numerical value which is the chain probability which is 3 group  $w_{m-2}w_{m-1}w_m$  of a word from appearing probability of word  $w_m$  under conditions to which  $c_m$  was given according to several 2 formulas.

[0098]B88: A block which adds a probability value calculated by B86 to the variable out.

[0099]B90: A block which outputs the variable out.

[0100]In this embodiment, when creating a word class, the single figure of the last of a code number of a classification lexical table was omitted and created, and an attached word was made into 1 word 1 class, but it is applicable also to a method of creating a word class by other methods. Although

a classification lexical table of a ward-facilities language research institute was used as a thesaurus, it cannot be overemphasized that it can apply when thesauri other than a classification lexical table are used. Only in  $n = 3$  (trigram) of a  $n$ -gram model, it illustrated, but of course, it can apply also to a value of other  $n$ .

[0101]An experimental result about example \*\* which generates a meaningless sentence using five sentences mentioned above is shown below. The result with which Perplexity compared a group of a generated meaningless sentence to descending (namely, order with small occurrence probability) is as follows :P By perplexity:260.36 motorcycle A cigarette. It has started. From this morning A building. It does not set. In a drawer A pigeon. It is in large numbers. On the head A key. It is painful. It is high. A gear tooth. It stopped.

[0102]Perplexity: By 260.36 motorcycles A cigarette. It stopped. From this morning A building. It does not set. In a drawer A pigeon. It is in large numbers. On the head A key. It is painful. It is high. A gear tooth. It has started.

[0103]Perplexity: By 260.27 motorcycles A cigarette. It has started. A building from this morning. It is painful. In a drawer A pigeon. It is in large numbers. On the head A key. It does not set. It is high. A gear tooth. It stopped.

[0104]Perplexity: By 260.27 motorcycles A cigarette. It stopped. From this morning A building is painful. In a drawer A pigeon. It is in large numbers. On the head A key. It does not set. It is high. A gear tooth. It has started.

[0105]Perplexity: By 258.92 motorcycles A pigeon. It is in large numbers. From this morning A building. It does not set. In a drawer A cigarette. It has started. On the head A key. It is painful. It is high. A gear tooth. It stopped.

[0106]Perplexity: By 258.92 motorcycles A pigeon. It is in large numbers. From this morning A building. It does not set. In a drawer A cigarette. It stopped. On the head A key. It is painful. It is high. A gear tooth. It has started.

[0107]Perplexity: By 258.83 motorcycles A pigeon. It is in large numbers. From this morning A building. It is painful. In a drawer A cigarette. It has started. On the head A key. It does not set. It is high. A gear tooth. It stopped.

[0108]Perplexity: By 258.83 motorcycles A pigeon. It is in large numbers. From this morning A building. It is painful. In a drawer A cigarette. It stopped. On the head A key does not set. It is high. A gear tooth. It has started.

[0109]On the other hand, a block of drawing 3 of B40 performed same processing using a word trigram (several 1 formula) for comparison. As a result, 16 sets of groups of a sentence (a probability value became the minimum) from which Perplexity became the maximum were obtained. These parts are shown below :P By perplexity:718.67 motorcycle A cigarette. It does not set. From this morning A gear tooth. It is painful. In a drawer A pigeon. It stopped. On the head A building. It has started. It is high. A key. It is in large numbers.

[0110]Perplexity: By 718.67 motorcycles A cigarette. It does not set. From this morning A gear tooth has started. In a drawer A pigeon. It stopped. On the head A building. It is painful. It is high. A key. It is in large numbers.

[0111]Perplexity: By 718.67 motorcycles A cigarette. It does not set. From this morning A gear tooth has started. In a drawer A pigeon. It stopped. On the head A building. It is in large numbers. It is high. A key. It is painful.

[0112]Perplexity: By 718.67 motorcycles A cigarette. It does not set. From this morning There are many gear teeth. In a drawer A pigeon. It stopped. On the head A building. It has started. It is high. A key. It is painful.

[0113]perplexity: By 718.67 motorcycles A cigarette. It stopped. From this morning A gear tooth has started. In a drawer A pigeon. It does not set. On the head A building. It is painful. It is high. A key. It is in large numbers.

[0114]Perplexity: By 718.67 motorcycles A cigarette. It stopped. From this morning A gear tooth has started. In a drawer A pigeon. It does not set. On the head A building. It is in large numbers. It is high. A key. It is painful.

[0115]Perplexity: By 718.67 motorcycles A cigarette. It stopped. From this morning There are many gear teeth. In a drawer A pigeon. It does not set. On the head A building. It has started. It is high. A key. It is painful.


[0116]When these experimental results were compared and a language processing method of this embodiment was used, by a case where a word trigram is used, a result that an owner meaning sentence of a basis was also contained was brought to being stabilized and constituting a meaningless sentence. This shows that this invention has sufficient effect.

[0117]Next, in order to investigate an effect of example \*\*, a case where back-off smoothing was used was compared with this invention by calculating Perplexity using a beginning sentence of actually broadcast news. Used data for evaluation is the text which it began was written with generating of an anchor in an existing news program on June 4, 1996, and a sentence which overlapped with a database for study is not contained. That Perplexity at the time of using back-off smoothing was [ for ] 62.38, when a language processing method of this embodiment was used, the result 57.97 was obtained, and the result checked an effect of this invention.

[0118]In this experiment, first, a class trigram model was adopted as a word trigram model and a class  $n$ -gram model as a word  $n$ -gram model, and perplexity compared only these both. Back-off smoothing of Katz was used together about a word trigram. A result is shown in Table 2. The word trigram showed perplexity lower than a class trigram from Table 2.

[0119]

[Table 2]

 ID=000009
---

[0120]In order to investigate this reason, a result which receives data for evaluation was examined in detail. As an example which has lowered perplexity of a word trigram model, a trigram shows an example used as a high value.

[0121]

[Equation 9]

$P(\text{the exercise} \mid \text{Pacific rim} \text{ --- congruent}) = 1P(\text{center} \text{ [ of council ] }, \text{education}) = 1$  -- in the example of these, the probability of the trigram is 1. That is, in learned data, when the chain of two say words appears [ the "Pacific rim" and / "" ], in the next, it means that the "exercise" always appeared. A "center", "education", and a "council" are the same. It seems that the value of word perplexity fell since the chain of such three words appeared also in the data for evaluation.

[0122]On the other hand, since each will be transposed to a class even if there is only "exercise" in learned data as a word which follows [ the "Pacific

rim" and / "" ] in the case of a class trigram model, probability of a class chain is not necessarily set to 1. Since appearing probability of an "exercise" in a class (1.305) which belongs can multiply probability of this class chain by an "exercise", a probability value of a class trigram model becomes still smaller. Perplexity of a class trigram model is considered to have become larger than a case where a word trigram model is used by such reason.

[0123]Conformity capability as opposed to learned data in a direction of :(1) word trigram model in which the following conclusions are obtained from the above examination is higher than a class trigram model. Therefore, when a statistical property of data for evaluation and data for study is alike, high expression ability is demonstrated.

[0124](2) A class trigram model has the same effect as smoothing in order to use a statistical property for every class. Therefore, although it cannot expect that a probability value is set to 1 like several 9 formulas, it is possible to presume the probability value also to a word chain which does not exist in learned data.

[0125]From the above (2), how to use a class trigram as a smoothing model for a word trigram can be considered. So, when 3 groups of a word which is not contained in a word trigram model appeared, not using back-off smoothing, a multiplicative model which presumes a probability value with a class trigram model was built, and it asked for perplexity to evaluation data using this model. As a result, to perplexity 62.38 in word trigram + back-off smoothing, the result 57.97 was obtained and it became possible to decrease perplexity further.

[0126]

[Effect of the Invention]As mentioned above, as explained, since a word is semantically classified according to a thesaurus, compared with the classification of a part of speech, also in the word according to which there were many kinds of a classification and it was classified, a degree of association becomes deep and the accuracy of chain probability becomes high as \*\* by the invention of claims 1 and 8.

[0127]Since the word itself is not used but the classification code of a classification with which the word is contained is used in the invention of claims 2 and 9 when acquiring the chain probability between words, For example, the number of combination of the classification code which memorizes the chain probability of the learned classification code decreases substantially compared with the combination of a word. Thereby, memory space of a language processor can be made smaller than before.

[0128]In the invention of claims 3 and 10, it is making \*\* of attached \*\*\*\* into a classification code about the attached word which is not indicated to a thesaurus, and handling becomes possible also about an attached word.

[0129]It is classifying a part of speech according to the invention of claims 4 and 11 about the independent word which is not indicated to the classification of the thesaurus, and the handling of many words becomes possible.

[0130]In the invention of claims 5 and 12, the meaningless sentence used in an acoustic-sense mental experiment can be created because chain probability chooses the word sequence used as the minimum, and the created sentence does not include an arbitrary element, either.

[0131]Since chain probability makes the word sequence used as the maximum the optimal candidate in voice recognition candidates in the invention of claims 6 and 13, compared with the speech recognition of only the conventional word, the chain between words is taken into consideration and speech recognition precision improves.

[0132]In the invention of claims 7 and 14, since the chain probability between words can be presumed even when the chain probability between words is not given by study etc., the word which is the target of handling can be made into a large number.

---

[Translation done.]

## \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

DESCRIPTION OF DRAWINGS

---

## [Brief Description of the Drawings]

- [Drawing 1] It is a block diagram showing the functional constitution of the language processor of a 1st embodiment of this invention.
- [Drawing 2] It is a flow chart which shows the procedure of the language processor of a 1st embodiment of this invention.
- [Drawing 3] It is a flow chart which shows the procedure of the language processor of a 1st embodiment of this invention.
- [Drawing 4] It is a block diagram showing the functional constitution of the language processor of a 2nd embodiment of this invention.
- [Drawing 5] It is a flow chart which shows the procedure of the language processor of a 2nd embodiment of this invention.
- [Drawing 6] It is a flow chart which shows the procedure of the language processor of a 2nd embodiment of this invention.

## [Description of Notations]

- 1 Meaningless sentence generation block
- 2 Morphological-analysis block
- 4 Thesaurus table storing block
- 6 Word classification block
- 8 The processing block which stores the table of the trigram probability between classes
- 10 The processing block which chooses the group of a meaningless sentence
- 22 Morphological-analysis block
- 24 Word trigram probability storing block
- 26 The processing block which reads word chain probability
- 28 Thesaurus storing block
- 30 Word classification block
- 32 The storing block which memorizes the table of the trigram probability between classes, and class word probability
- 34 The processing block which calculates a class trigram
- 36 The processing block which chooses a recognition candidate using class trigram probability

---

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-110385

(43) 公開日 平成11年(1999) 4月23日

(51) Int.Cl.<sup>6</sup>

G 0 6 F 17/27

識別記号

F I

G 0 6 F 15/38

E

審査請求 未請求 請求項の数14 O L (全 15 頁)

(21) 出願番号 特願平9-268592

(22) 出願日 平成9年(1997)10月1日

特許法第30条第1項適用申請有り 1997年6月20日 社団法人電子情報通信学会主催の「電子情報通信学会技術研究報告」において文書をもって発表

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

(72) 発明者 安藤 彰男

東京都世田谷区砧一丁目10番11号 日本放送協会 放送技術研究所内

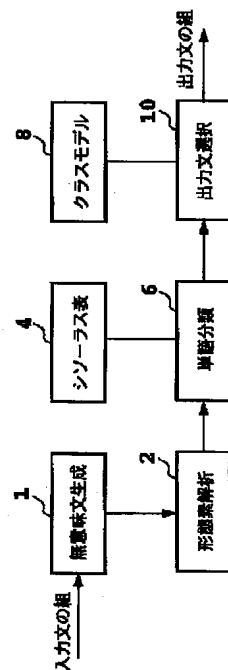
(74) 代理人 弁理士 谷 義一 (外3名)

(54) 【発明の名称】 言語処理装置および方法

(57) 【要約】

【課題】 処理に要するメモリの容量を減じる。

【解決手段】 単語分類ブロック6はシソーラスにしたがって、形態素解析ブロック2にから出力される単語を分類する。



## 【特許請求の範囲】

【請求項1】 1つのクラスが複数の単語の集合で構成される複数のクラスに、複数の単語を分類し、当該分類されたクラス間の連鎖の程度を示す連鎖確率を取得し、当該取得された連鎖確率に基づきクラスおよび／または単語列を選択する言語処理装置において、シソーラスの分類を記憶する記憶手段と、該シソーラスの分類にしたがって、前記複数の単語を分類する分類手段とを具えたことを特徴とする言語処理装置。

【請求項2】 請求項1に記載の言語処理装置において、前記分類手段は、前記複数の単語を分類する時に、単語を分類コードに変換し、前記連鎖確率の計算時には前記単語列を分類コード列の形態で計算することを特徴とする言語処理装置。

【請求項3】 請求項2に記載の言語処理装置において、前記分類手段は、前記単語に付属する付属語については該付属語そのものを分類コードとすることを特徴とする言語処理装置。

【請求項4】 請求項2に記載の言語処理装置において、前記分類手段は、前記単語が前記シソーラスの分類に記載されていない自立語の場合には、品詞ごとの分類を行うことを特徴とする言語処理装置。

【請求項5】 請求項1に記載の言語処理装置において、前記連鎖確率が最小となる単語列を聴覚実験で使用する無意味文として選択する選択手段をさらに具えたことを特徴とする言語処理装置。

【請求項6】 請求項1に記載の言語処理装置において、前記連鎖確率が最大となる単語列を音声認識候補の中の最適候補として選択する選択手段をさらに具えたことを特徴とする言語処理装置。

【請求項7】 請求項1に記載の言語処理装置において、各単語間の連鎖確率の計算に関連して、クラスの中の単語の出現確率およびクラス間の連鎖確率を取得しておき、単語間の連鎖確率が直接取得できない場合には、前記単語の出現確率および前記クラス間の連鎖確率に基づき前記単語間の連鎖確率を推定することを特徴とする言語処理装置。

【請求項8】 1つのクラスが複数の単語の集合で構成される複数のクラスに、複数の単語を分類し、当該分類されたクラス間の連鎖の程度を示す連鎖確率を取得し、当該取得された連鎖確率に基づきクラスおよび／または単語列を選択する言語処理方法において、シソーラスの分類を記憶しておき、該シソーラスの分類にしたがって、前記複数の単語を分類することを特徴とする言語処理方法。

【請求項9】 請求項8に記載の言語処理方法において、前記複数の単語を分類する時に、単語を分類コードに変換し、前記連鎖確率の計算時には前記単語列を分類コード列の形態で計算することを特徴とする言語処理方

法。

【請求項10】 請求項9に記載の言語処理方法において、前記単語に付属する付属語については該付属語そのものを分類コードとすることを特徴とする言語処理方法。

【請求項11】 請求項9に記載の言語処理方法において、前記単語が前記シソーラスの分類に記載されていない自立語の場合には、品詞ごとの分類を行うことを特徴とする言語処理方法。

10 【請求項12】 請求項8に記載の言語処理方法において、前記連鎖確率が最小となる単語列を聴覚実験で使用する無意味文として選択することを特徴とする言語処理方法。

【請求項13】 請求項8に記載の言語処理方法において、前記連鎖確率が最大となる単語列を音声認識候補の中の最適候補として選択することを特徴とする言語処理方法。

【請求項14】 請求項8に記載の言語処理方法において、各単語間の連鎖確率の計算に関連して、クラスの中の単語の出現確率およびクラス間の連鎖確率を取得しておき、単語間の連鎖確率が直接取得できない場合には、前記単語の出現確率および前記クラス間の連鎖確率に基づき前記単語間の連鎖確率を推定することを特徴とする言語処理方法。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、自然言語処理において、最適な単語列を選択／生成する言語処理装置および方法に係り、特に、シソーラスに基づいて設定された単語クラス間の統計情報を用いる言語処理装置および方法に関する。

【0002】

【従来の技術】従来は、 $n$ 個の単語の統計的連鎖情報を利用する単語 $n$ -gramモデルが言語処理装置において利用されていた。また、単語を品詞ごとに分類して単語クラスを作成し、これらの単語クラス間（品詞間）の統計連鎖情報を利用するクラス $n$ -gramモデルも利用されていた（以上については、例えば、北他「音声言語処理」森北出版参照）。これらの方式では、例えば、単語 $n$ -gramモデルの場合、扱う単語数が増え、 $n$ が大きくなると、単語間の連鎖確率を十分に推定できなくなる、いわゆるスパース問題と呼ばれる問題が生じた。一方、品詞間のクラス $n$ -gramモデルの場合には、大ざっぱな言語処理しか行えないという欠点があった。

【0003】

【発明が解決しようとする課題】聴覚心理実験などにおいて、被験者が音声をきちんと聞き取れるかどうかを判定する実験が行われるが、この際、意味的な類推の影響を除去して本来の聴覚的特性のみを調べたいという要求

がある。このような目的を満足させる方法として、文法的に正しいが意味の通らない文を作成する方法がある

（例えば、渡部一雄：“話速交換による話速遅延の高齢難聴者に対する効果”、日本耳鼻咽喉科学会会報、Vol. 99, pp. 445-453 (1996)）。しかしながら、意味の通らない試験文を作成する良い方法が見つかっておらず、作成された試験文から、作成者の恣意的要素を除くことは事実上不可能であった。

【0004】一方、音声認識では、認識性能を向上させるため、音響レベルでの認識結果から言語的に尤もらしい単語列を選択したり、それまで得られた認識結果を用いて、それ以降の認識結果として想定しうる候補の数を削減するため、単語n-gramが利用されている。特に、大語彙連続音声認識では、単語n-gramの利用は不可欠とされている。また、音声認識以外の自然言語処理においても、複数個ある単語列の中から、尤もらしい単語列を選択するため、単語n-gramが利用されている。ところが、単語n-gramモデルでは、例えばn=3（3単語の連鎖モデル：トライグラムモデル）の場合を例にとると、単語数が2万語の場合には、2万の3乗（8,000,000,000,000）個の組み合わせを考える必要があり、処理の際に膨大なメモリが必要となるほか、入手可能な言語データベースからでは、その統計量を推定できないという問題がある（この問題は、スパースネス問題と呼ばれている）。この問題は、nが3より大きい場合には一層顕著となる。このような問題が、音声認識などでは、処理結果に大きな影響を与えていた。

【0005】そこで、本発明の目的は、単語を品詞毎に分類するよりも単語間の連鎖確率の精度を高め、また、言語処理装置のメモリに関する制限を緩和することの可能な言語処理装置および方法を提供することにある。

【0006】

【課題を解決するための手段】このような目的を達成するために、請求項1の発明は、1つのクラスが複数の単語の集合で構成される複数のクラスに、複数の単語を分類し、当該分類されたクラス間の連鎖の程度を示す連鎖確率を取得し、当該取得された連鎖確率に基づきクラスおよび／または単語列を選択する言語処理装置において、シソーラスの分類を記憶する記憶手段と、該シソーラスの分類にしたがって、前記複数の単語を分類する分類手段とを具えたことを特徴とする。

【0007】請求項2の発明は、請求項1に記載の言語処理装置において、前記分類手段は、前記複数の単語を分類する時に、単語を分類コードに変換し、前記連鎖確率の計算時には前記単語列を分類コード列の形態で計算することを特徴とする。

【0008】請求項3の発明は、請求項2に記載の言語処理装置において、前記分類手段は、前記単語に付属する付属語については該付属語そのものを分類コードとす

ることを特徴とする。

【0009】請求項4の発明は、請求項2に記載の言語処理装置において、前記分類手段は、前記単語が前記シソーラスの分類に記載されていない自立語の場合には、品詞ごとの分類を行うことを特徴とする。

【0010】請求項5の発明は、請求項1に記載の言語処理装置において、前記連鎖確率が最小となる単語列を聴覚実験で使用する無意味文として選択する選択手段をさらに具えたことを特徴とする。

10 【0011】請求項6の発明は、請求項1に記載の言語処理装置において、前記連鎖確率が最大となる単語列を音声認識候補の中の最適候補として選択する選択手段をさらに具えたことを特徴とする。

【0012】請求項7の発明は、請求項1に記載の言語処理装置において、各単語間の連鎖確率の計算に関連して、クラスの中の単語の出現確率およびクラス間の連鎖確率を取得しておき、単語間の連鎖確率が直接取得できない場合には、前記単語の出現確率および前記クラス間の連鎖確率に基づき前記単語間の連鎖確率を推定することを特徴とする。

20 【0013】請求項8の発明は、1つのクラスが複数の単語の集合で構成される複数のクラスに、複数の単語を分類し、当該分類されたクラス間の連鎖の程度を示す連鎖確率を取得し、当該取得された連鎖確率に基づきクラスおよび／または単語列を選択する言語処理方法において、シソーラスの分類を記憶しておき、該シソーラスの分類にしたがって、前記複数の単語を分類することを特徴とする。

30 【0014】請求項9の発明は、請求項8に記載の言語処理方法において、前記複数の単語を分類する時に、単語を分類コードに変換し、前記連鎖確率の計算時には前記単語列を分類コード列の形態で計算することを特徴とする。

【0015】請求項10の発明は、請求項9に記載の言語処理方法において、前記単語に付属する付属語については該付属語そのものを分類コードとすることを特徴とする。

40 【0016】請求項11の発明は、請求項9に記載の言語処理方法において、前記単語が前記シソーラスの分類に記載されていない自立語の場合には、品詞ごとの分類を行うことを特徴とする。

【0017】請求項12の発明は、請求項8に記載の言語処理方法において、前記連鎖確率が最小となる単語列を聴覚実験で使用する無意味文として選択することを特徴とする。

【0018】請求項13の発明は、請求項8に記載の言語処理方法において、前記連鎖確率が最大となる単語列を音声認識候補の中の最適候補として選択することを特徴とする。

50 【0019】請求項14の発明は、請求項8に記載の言

語処理方法において、各単語間の連鎖確率の計算に関連して、クラスの中の単語の出現確率およびクラス間の連鎖確率を取得しておき、単語間の連鎖確率が直接取得できない場合には、前記単語の出現確率および前記クラス間の連鎖確率に基づき前記単語間の連鎖確率を推定することを特徴とする。

【0020】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を詳細に説明する。

$$P(w_m | w_1 w_2 \dots w_{m-1})$$

$$\approx P(w_m | w_{m-n+1} w_{m-n+2} \dots w_{m-1})$$

【0023】で近似する(ただし $m \geq n$ )。今、各単語は、それぞれ1つのクラスにしか属さないものとする。

単語 $w_i$ が属するクラスを $c_i$ で表す。このとき、クラス $n$ -gramモデルでは、単語列 $w_1 w_2 \dots w_{m-1}$ が

$$P(w_m | w_1 w_2 \dots w_{m-1})$$

$$\approx (P(c_m | c_{m-n+1} c_{m-n+2} \dots c_{m-1}) \cdot P(w_m | c_m))$$

【0025】また、聴覚心理実験のための試験文生成に当たっては、数2式を簡略化した式：

【0026】

【数3】 $P(c_m | c_{m-n+1} c_{m-n+2} \dots c_{m-1})$

も利用する。

【0027】クラス $n$ -gramモデルでは、単語をクラスに分類し、クラス間の連鎖の程度を示す確率(連鎖確率)と、クラス内の単語の出現確率を用いるものであり、単語 $n$ -gramモデルと比べて、パラメータ数を★

単語 $n$ -gramモデルとクラス $n$ -gramモデルとの独立パラメータ数の比較

n	1	2	3
単語 $n$ -gramモデル	19,999	399,999,999	7,999,999,999,999
クラス $n$ -gramモデル	19,999	1,018,999	1,000,018,999

(単語数20,000、クラス数1,000の場合)

【0029】クラス $n$ -gramモデルとしては、従来から品詞に基づいて単語クラスを設定する方法などが提案されているが、大ざっぱな単語分類しかできないという問題があった。また、単語間の相互情報量を用いて単語を分類する方法も提案されているが、最適な分類を求めるアルゴリズムが存在しないなどの問題が知られている。一方、単語の意味的な分類を行うための手法として、シソーラスが知られている。そこで、実施の形態では、このシソーラスに基づいたクラス $n$ -gramモデルを利用することにより、スパースネス問題など、単語 $n$ -gramモデルにおける問題の解決を図る。

【0030】このために、

\*【0021】まず、本発明を適用した言語処理方法について述べる。本実施形態では、シソーラスに基づくクラス $n$ -gramモデルを利用する。単語 $n$ -gramモデルは、単語列の生起を $n-1$ 重マルコフ課程で近似するモデルである。従って、単語列 $w_1 w_2 \dots w_{m-1}$ が与えられた時、単語 $w_m$ が生起する確率を、式

【0022】

【数1】

\*与えられたという条件のもとで、単語 $w_m$ が生起する確率を、次式で近似する：

【0024】

【数2】

20★大幅に減らすことが可能である。例えば、扱う語彙の大きさ(単語数)を $V$ 、クラスの数 $C$ とすると、単語 $n$ -gramにおける独立したパラメータの数は $V^n - 1$ である。一方、クラス $n$ -gramモデルの独立パラメータの数は、 $C^n - 1 + V - C$ となる。 $V=20,000$ 、 $C=1,000$ とした場合の独立パラメータの数を、表1に示す。

【0028】

【表1】

①聴覚心理実験のための聴き取り検査用の無意味文を生成させた例(第1実施形態)

②音声認識のための単語 $n$ -gramモデルにおけるスパースネス問題への対処例(第2実施形態)の2つの場合を例示する。なお、いずれの場合も、 $n$ -gramの $n$ を3とした場合(トライグラム(三つ組の連鎖確率))について示す。

【0031】なお、第1および第2実施形態では、シソーラスとして、国立国語研究所の分類語彙表を利用した例を示す。分類語彙表では、各単語のコード番号として、1.3062など、ピリオドで区切られた5桁の数字が与えられている。このコード番号をそのまま1つの

クラスを表す数字とすると、クラスの数が大きくなりすぎるため、本実施形態では、最後の1桁を切り捨てた後、同じコード番号を有する単語の集合を、一つのクラスとして扱う。

【0032】分類語彙表には、基本的に自立語しか登録されていないため、数2式のクラスモデルを利用する際には、付属語の取り扱いが問題となる。また、分類語彙表に登録されていない単語をどう扱うかも問題となる。付属語については、文法的に分類することによりクラスを構成することも可能であるが、第1および第2実施形態では、1語で1クラスを構成させる方法を採用した。また、分類語彙表に登録されていない単語については、クラスモデル学習時に、形態素解析結果に基づいて自動的に登録する方法を採用した。

【0033】単語n-gramの学習は、通常、言語データベース中の文を、形態素解析によって単語単位に分割した後、単語連鎖の頻度から数1式の確率値を推定する。クラスn-gramも学習も同様の手順で行う。

【0034】まず、言語データベースを単語単位に分割し、分類語彙表と照合することにより、単語をクラスコードに変換する。その際、付属語は、コードへの変換を行わず、語そのものをクラスコードとする。分類語彙表に記載されていない自立語については、以下の処理を行う。まず動詞などの活用形については、語幹で照会した後、その活用形と対応する分類コード番号を別に用意する単語-クラス対応表に追加する。数詞は、強制的に数詞のコードを割り当てた後、その数詞も単語-クラス対応表に追加する。名詞については、形態素解析結果に従って、「普通名詞」、「サ変名詞」、「固有名詞」、「形式名詞」、「副詞の名詞」、「時相名詞」のクラスに振り分け、合わせて単語-クラス対応表（クラスとそのクラスに含まれる単語の対応関係を示す表）への追加を行う。それ以外の語については、未知クラスに分類する。

【0035】単語トライグラム、クラストライグラムの学習には、放送局のニュース原稿データベースを利用した。これは、1992年から1996年の5月までに、放送局の記者が作成したニュース原稿をデータベース化\*

$$H(L) = - \sum_{w_1 \dots w_n} \frac{1}{n} P(w_1 \dots w_n) \log P(w_1 \dots w_n)$$

【0041】としたとき、

【0042】

【数5】 $PP = 2^{H(L)}$

で表される量である。従って、文の生起確率が小さいほど、Perplexityは大きくなる。

【0043】このような言語処理方法で、聴き取り検査用の無意味文生成のための言語処理を行う言語処理装置の機能構成を示す。図1において、1は入力文の組（先に示した有意味文の5文）をそれぞれ3つの部分に分割

\*したものである。クラストライグラムの学習では、語彙の大きさを20,000語とし、それ以外の単語は、未知語専用のクラスに割り当てた。この未知語専用クラスは、未知クラスとは異なるものである（前者は、語彙に含まれない単語のクラスであり、後者は語彙の範囲に含まれる単語であるが、属するクラスが未知の単語の集合である）。得られたクラスの数 は 936 であり、そのうち518クラスは分類語彙表から作成された。自立語に対する各クラスに割り当てられた平均単語数は41であった。一方、単語トライグラムについても同じ条件で作成する。

【0036】（第1実施形態）以下、例①について、詳しく説明する。音声の聴き取り試験を行う際、確実に聞き取られたどうかを調査するためには、平易な単語からなる意味不明な文を作成し、この文を読み上げた音声を用いて実験を行うことが考えられる。以下、このような無意味文の生成について例示する。

【0037】無意味文の作成は、あらかじめ用意された短い有意味文をそれぞれ3つの部分に分割し、これらの部分を入れ換えることにより行う。有意味文の例を以下に示す。

【0038】オートバイでは 煙草が すえない。今朝から 歯が 痛い。引き出しに 鍵が かかっている。頭に 鳩が 止った。高い ビルが たくさんある。

【0039】ただし簡便化のため、各文を、分類語彙表および単語-クラス対応表を用いてクラスコード列に変換した後、数3式のクラス間のトライグラム確率のみを用いて、文ごとの生起確率を計算し、最も確率が低くなった文の組を選択することにより、無意味文を作成する。また、本発明の効果を比較により示すため、単語トライグラムを用いて同様な処理を行った。なお、以下では、便宜上、確率値のかわりに、Perplexityを利用して、結果を表示する。この場合のPerplexityは、各文（単語列 $w_1 \dots w_n$ で表す）の生起確率を $P(w_1 \dots w_n)$ 、そのエントロピーを

【0040】

【数4】

し、これらを入れ換えることにより無意味文を生成する処理ブロックである。この場合に生成される文の数は14,400である。

【0044】2は、生成された無意味文の組を単語単位に分割する形態素解析ブロックである。4はシソーラスを格納する格納ブロックである。6はシソーラスに基づいて単語をクラスコードに変換する単語分類ブロックである。

【0045】また、8はクラス間のトライグラム確率

(数3式の確率値)の表を記憶する格納ブロックである。10は、生成された全ての無意味文の組についてクラストライグラムに基づいてPerplexityを計算し、最もPerplexityが大きくなった無意味文の組を選択する処理ブロックである。

【0046】図2は言語処理装置において実行する言語処理の内容を示す。図2に従って、実際の言語処理を説明する。図2の各ブロックにおける実行内容は以下のとおりである。

【0047】B2：入力文の組を読み込むブロック。

【0048】B4：各入力文を3つの部分に分割し、その第1の部分の固定し、第2、第3の部分を入れ換えることにより、無意味文を生成するブロック。5つの文を、

【0049】

【数6】

$b_{11} b_{12} b_{13}$

$b_{21} b_{22} b_{23}$

$b_{31} b_{32} b_{33}$

$b_{41} b_{42} b_{43}$

$b_{51} b_{52} b_{53}$

で表す( $b_{ij}$ は*i*番目の文の*j*番目のブロック)。また、関数*r*を、集合{1, 2, 3, 4, 5}から{1, 2, 3, 4, 5}への関数で、 $i \neq j \Rightarrow r(i) \neq r(j)$ を満たすものとする。また、関数*s*も同様に定義する。この時、本ブロックでは、

【0050】

【数7】

$b_{11} b_{r,2} b_{s,3}$

$b_{21} b_{r,2} b_{s,3}$

$b_{31} b_{r,2} b_{s,3}$

$b_{41} b_{r,2} b_{s,3}$

$b_{51} b_{r,2} b_{s,3}$

【0051】で表される文を生成する。ここに、 $r_i$ 、 $s_i$ は、それぞれ、 $r(i)$ 、 $s(i)$ の略記である。

【0052】B6：変数Maxに0を代入するブロック。

【0053】B8：B4で生成された全ての無意味文の組が選択されたかどうかを判断するブロック。

【0054】B10：B4で生成された無意味文の組の中から、まだ選択されていない組を選択するブロック。

【0055】B12：B10で選択された無意味文の組に対して、そのPerplexityを計算するサブルーチン。詳細は、図3に示す。入力は無意味文の組で、計算結果は、変数outに返される。

【0056】B14：変数outの値が、変数Maxより大きいことを判定するブロック。

【0057】B16：変数outの値が、変数Maxと等しいことを判定するブロック。

【0058】B18：出力文の組を格納する配列を初期化し、そこに、現在選択されている無意味文の組を格納するブロック。

【0059】B20：変数Maxの値を、変数outの値で置き換えるブロック。

【0060】B22：出力文の組を格納する配列に、現在選択されている無意味文の組を追加するブロック。

【0061】B24：格納されている出力文の組を出力するブロック。

【0062】また、図3は、図2のB12に示したサブルーチンでの処理内容を示す。図3の処理ブロックは以下のブロックにより構成されている。

【0063】B32：変数outに0を代入するブロック。

【0064】B34：入力された無意味文の組から、文を選択する際、全ての文が選択されたことを判定するブロック。

【0065】B36：入力された文の組の中から、まだ選択されていない文を選択するブロック。

【0066】B37：選択された文を単語単位に分割する形態素解析ブロック。形態素解析法としては、どのような方法を用いてもよいが、本実施形態では、京都大学で開発されたソフトウェアJumanを採用した。

【0067】B38：文を単語の3つ組に分解するブロック。文が単語 $w_1, w_2, \dots, w_n$ の連結でできている場合、3つ組 $w_1 w_2 w_3, w_2 w_3 w_4, \dots, w_{n-2} w_{n-1} w_n$ を生成する。

【0068】B40：単語3つ組中の各単語をクラスに分類し、クラスコードの3つの組に変換するブロック。単語3つの組 $w_1 w_2 w_3, w_2 w_3 w_4, \dots, w_{n-2} w_{n-1} w_n$ を、クラス3つ組 $c_1 c_2 c_3, c_2 c_3 c_4, \dots, c_{n-2} c_{n-1} c_n$ に変換する。

【0069】B41：B40で生成された各3つ組について、数3式のトライグラム確率を読み出すブロック。

【0070】B42：B40で読み出したトライグラム確率から、式

【0071】

【数8】

$$P(c_1 \cdots c_n) = P(c_3 | c_1, c_2) \cdot P(c_4 | c_2, c_3) \cdots P(c_n | c_{n-2}, c_{n-1})$$

$$H(L) = - \sum_{w_1 \cdots w_n} \frac{1}{n} P(c_1 \cdots c_n) \log P(c_1 \cdots c_n)$$

$$PP = 2^{H(L)}$$

【0072】によって文のPerplexityを計算するブロック。

【0073】B44：変数outに、変数outの値とB42で求めたPerplexityの値の和を代入するブロック。

【0074】B46：変数outの値を出力するブロック。

【0075】(第2実施形態)次に例②について、以下に詳しく説明する。一般に、音声認識で単語トライグラムモデルを利用する際、発声された文の中に含まれる単語の3つ組が、学習時に存在しなかったため、その確率値が推定できていない場合がある。例②では、このよう

な場合に対して、数2式のクラストライグラムを用いて、その確率値を推定する。  
【0076】図4は例②の言語処理を実行する言語処理装置の機能構成を示す。この例では、なんらかの連続音声認識システムがあって、このシステムが、複数の認識候補文を出力する場合を想定している。図4において、22は、音声認識候補として入力された複数の文を、単語単位に分割する形態素解析ブロック、24は、単語トライグラム確率(数1式)を記憶する格納ブロックである。26は、入力に対応する単語連鎖確率を読み出す処理ブロックである。28はシソーラスを格納する格納ブロック、30はシソーラスに基づいて単語をクラスコードに変換する単語分類ブロックである。

【0077】また、32はクラス間のトライグラム確率と、クラス—単語確率(単語がそのクラスに出現する確率)の表を記憶する格納ブロックである。34は、入力文に対応するクラストライグラム確率(数2式の確率)を計算する処理ブロックである。36は、単語トライグラム確率が求められていなかった場合にはクラストライグラム確率を利用しながら、入力された文(認識候補)に対するPerplexityを計算し、その値が最も小さかったもの(確率値が高かったもの)を選択して出力する処理ブロックである。

【0078】図5は、例2における言語処理の内容を示す。図5に従って、実際の言語処理を説明する。図5の各ブロックにおける実行内容は以下の通りである。

【0079】B52：入力文の組(複数の認識候補)を読み込むブロック。

【0080】B54：変数Minの値に、可能なかぎり小さい値を代入するブロック。

【0081】B56：入力された文を全て選択されたかどうかを判定するブロック。

【0082】B58：それまでに選択されていない文を選択するブロック。

【0083】B60：選択された文に対するPerplexityを計算するサブルーチン。入力は、選択された文であり、それに対するPerplexityを変数outを用いて返す。

【0084】B62：変数outが、変数Minより小さいかどうかを判定するブロック。

【0085】B64：出力文を格納する配列を初期化し、その配列に、現在選択されている文を格納する。

【0086】B66：変数Minに変数outの値を代入する。

【0087】B68：出力文用の配列に格納されている文を出力する。

【0088】図6は、サブルーチンB60の処理内容を示す。図6の処理ブロックは以下のブロックで構成されている。

【0089】B72：変数outに0を代入するブロック。

【0090】B73：入力された文を単語単位に分割する形態素解析ブロック。

【0091】B74：文を単語の3つ組に分解するブロック。文が単語 $w_1, w_2, \dots, w_n$ の連結でできている場合、単語の3つ組 $w_1 w_2 w_3, w_2 w_3 w_4, \dots, w_{n-2} w_{n-1} w_n$ を生成する。

【0092】B76：全ての単語の3つ組が選択されたかどうかを判定するブロック。

【0093】B78：それまでに選択されていない単語3つ組を選択するブロック。

【0094】B80：選択された3つ組に対し、単語トライグラムが存在するかどうかを判定するブロック。

【0095】B82：単語トライグラムモデルから求められた確率の対数値を、変数outに加算するブロック。

【0096】B84：選択されている単語の3つ組から、クラスの3つ組を求める単語分類ブロック。

【0097】B86：単語の3つ組を $w_{n-2}, w_{n-1}, w_n$ 、単語 $w_i$ の属する単語クラスを $c_i$ とするとき、クラスの3つ組 $c_{n-2}, c_{n-1}, c_n$ に対する連鎖確率と、 $c_n$ が与えられた条件下での単語 $w_n$ の出現確率から、数

2式に従って単語の3つ組 $w_{n-2}, w_{n-1}, w_n$ の連鎖確率の対数値を求めるブロック。

【0098】B88: B86で求めた確率値を変数outに加算するブロック。

【0099】B90: 変数outを出力するブロック。

【0100】本実施形態では、単語クラスを作成する際、分類語彙表のコード番号の最後の1桁を切り捨て作成し、また付属語は1単語1クラスとしたが、その他の方法で単語クラスを作成する方法にも適用可能である。さらに、シソーラスとして区立国語研究所の分類語彙表を用いたが、分類語彙表以外のシソーラスを利用した場合に適用可能なことはいうまでもない。また、n-gramモデルの $n=3$ の場合(トライグラム)の場合にのみ例示したが、その他の $n$ の値に対しても適用可能であることは勿論である。

【0101】上述した5つの文を用いて無意味文を生成する例①についての実験結果を以下に示す。生成された無意味文の組を、Perplexityが大きい順(すなわち生起確率が小さい順)に並べた結果は、以下の通りである:  
Perplexity: 260.36

オートバイでは 煙草が かかっている。今朝から ビルが すえない。引き出しに 鳩が たくさんある。頭に 鍵が 痛い。高い 歯が 止った。

【0102】Perplexity: 260.36

オートバイでは 煙草が 止った。今朝から ビルが すえない。引き出しに 鳩が たくさんある。頭に 鍵が 痛い。高い 歯が かかっている。

【0103】Perplexity: 260.27

オートバイでは 煙草が かかっている。今朝から ビルが 痛い。引き出しに 鳩が たくさんある。頭に 鍵が すえない。高い 歯が 止った。

【0104】Perplexity: 260.27

オートバイでは 煙草が 止った。今朝から ビルが 痛い。引き出しに 鳩が たくさんある。頭に 鍵が すえない。高い 歯が かかっている。

【0105】Perplexity: 258.92

オートバイでは 鳩が たくさんある。今朝から ビルが すえない。引き出しに 煙草が かかっている。頭に 鍵が 痛い。高い 歯が 止った。

【0106】Perplexity: 258.92

オートバイでは 鳩が たくさんある。今朝から ビルが すえない。引き出しに 煙草が 止った。頭に 鍵が 痛い。高い 歯が かかっている。

【0107】Perplexity: 258.83

オートバイでは 鳩が たくさんある。今朝から ビルが 痛い。引き出しに 煙草が かかっている。頭に 鍵が すえない。高い 歯が 止った。

【0108】Perplexity: 258.83

オートバイでは 鳩が たくさんある。今朝から ビルが 痛い。引き出しに 煙草が 止った。頭に 鍵が

すえない。高い 歯が かかっている。

【0109】一方、比較のため、図3のB40のブロックで、単語トライグラム(数1式)を用いて、同様の処理を行った。その結果、Perplexityが最大となった(確率値が最小となった)文の組が、16組得られた。以下にこれらの一部を示す:

Perplexity: 718.67

オートバイでは 煙草が すえない。今朝から 歯が 痛い。引き出しに 鳩が 止った。頭に ビルが かかっている。高い 鍵が たくさんある。

【0110】Perplexity: 718.67

オートバイでは 煙草が すえない。今朝から 歯が かかっている。引き出しに 鳩が 止った。頭に ビルが 痛い。高い 鍵が たくさんある。

【0111】Perplexity: 718.67

オートバイでは 煙草が すえない。今朝から 歯が かかっている。引き出しに 鳩が 止った。頭に ビルが たくさんある。高い 鍵が 痛い。

【0112】Perplexity: 718.67

20 オートバイでは 煙草が すえない。今朝から 歯が たくさんある。引き出しに 鳩が 止った。頭に ビルが かかっている。高い 鍵が 痛い。

【0113】perplexity: 718.67

オートバイでは 煙草が 止った。今朝から 歯が かかっている。引き出しに 鳩が すえない。頭に ビルが 痛い。高い 鍵が たくさんある。

【0114】Perplexity: 718.67

30 オートバイでは 煙草が 止った。今朝から 歯が かかっている。引き出しに 鳩が すえない。頭に ビルが たくさんある。高い 鍵が 痛い。

【0115】Perplexity: 718.67

オートバイでは 煙草が 止った。今朝から 歯が たくさんある。引き出しに 鳩が すえない。頭に ビルが かかっている。高い 鍵が 痛い。

【0116】これらの実験結果を比較すると、本実施形態の言語処理方法を用いた場合は、無意味文を安定して構成しているのに対し、単語トライグラムを利用した場合では、もとの意味文も含まれるという結果となった。このことより、本発明が、十分な効果を有することが分かる。

40 【0117】次に、例②の効果を調べるため、実際に放送されたニュースの書き起し文を用いてPerplexityを計算することにより、バックオフ平滑化を用いた場合と本発明とを比較した。用いた評価用データは、1996年6月4日のあるニュース番組におけるアンカーの発生から書き起したテキストであり、学習用のデータベースと重複した文は含まれていない。その結果、バックオフ平滑化を用いた場合のPerplexityが62.38であったのに対して、本実施形態の言語処理方法を用いた場合は57.97という結果が得られ、本発明の効果が確認され

た。

【0118】この実験ではまず、単語n-gramモデルとして単語トライグラムモデル、クラスn-gramモデルとしてクラストライグラムモデルを採用し、パーブレキシティによりこの両者のみの比較を行った。なお、単語トライグラムについては、Katzのバックオフ平\*

パーブレキシティによる言語モデルの比較

単語モデル	パーブレキシティ
単語トライグラムモデル 式(1)のモデル	62.38
クラストライグラムモデル 式(2)のモデル	186.16
クラス間トライグラム 式(3)のモデル	33.76

【0120】この理由を調べるため、評価用のデータに対する結果を詳細に検討した。単語トライグラムモデルのパーブレキシティを下げている例として、トライグラムが高い値となった例を示す。

【0121】

【数9】

P(演習|環太平洋、合同)=1

P(審議会|中央、教育)=1

これらの例では、トライグラムの確率が1となっている。すなわち、学習データの中では、「環太平洋」、「合同」という2単語の連鎖が現れた場合には、その次には必ず「演習」が現れたということを意味する。「中央」、「教育」、「審議会」も同様である。このような3単語の連鎖が評価用データの中にも出現したため、単語パーブレキシティの値が下がったものと思われる。

【0122】これに対して、クラストライグラムモデルの場合は、たとえ、学習データにおいて、「環太平洋」、「合同」の後に続く単語としては「演習」しかなかったとしても、それぞれがクラスに置き換えられるため、クラス連鎖の確率は、必ずしも1にはならない。さらに、このクラス連鎖の確率に、「演習」が属するクラス(1, 305)における「演習」の出現確率が乗じられるので、クラストライグラムモデルの確率値は、一層小さくなる。このような理由によって、クラストライグラムモデルのパーブレキシティが、単語トライグラムモデルを利用した場合よりも大きくなったものと思われる。

【0123】以上の検討より、次のような結論が得られる：

(1) 単語トライグラムモデルの方が、クラストライグラムモデルよりも、学習データに対する適合能力が高い。従って、評価用データと学習用データの統計的性質が似通っている場合には、高い表現能力を発揮する。

\*滑化を併用した。結果を表2に示す。表2より、単語トライグラムの方が、クラストライグラムよりも低いパーブレキシティを示した。

【0119】

【表2】

【0124】(2)クラストライグラムモデルは、クラスごとの統計的性質を利用するため、スムージングと同様の効果を有する。従って、数9式のように確率値が1になることは望めないが、学習データ中に存在しない単語連鎖に対しても、その確率値を推定することが可能である。

【0125】上記(2)より、クラストライグラムを単語トライグラムのためのスムージングモデルとして利用する方法が考えられる。そこで、単語トライグラムモデルに含まれていない単語の3つ組が出現した際には、バックオフスムージングを用いず、クラストライグラムモデルによって確率値を推定する複合モデルを構築し、このモデルを用いて、評価データに対するパーブレキシティを求めた。その結果、単語トライグラム+バックオフ平滑化の場合のパーブレキシティ62.38に対して、57.97という結果が得られ、パーブレキシティをさらに減少させることが可能となった。

【0126】

【発明の効果】以上、説明したように、請求項1、8の発明では、シソーラスにしたがって、単語を意味的に分類するので、品詞の分類に比べ分類の種類数が多く、分類された単語も関連度が深くなり、以って、連鎖確率の精度が高くなる。

【0127】請求項2、9の発明では、単語間の連鎖確率を取得する際に単語そのものを使用せず、その単語が含まれる分類の分類コードを使用するので、たとえば、学習した分類コードの連鎖確率を記憶しておく分類コードの組み合わせ数は、単語の組み合わせに比べると大幅に減少する。これにより、言語処理装置のメモリ容量は従来よりも小さくすることができる。

【0128】請求項3、10の発明では、シソーラスに記載されていない付属語については付属語そのものを分類コードとすることで、付属語についても取り扱いが可能

となる。

【0129】請求項4、11の発明では、シソーラスの分類に記載されていない自立語については品詞の分類を行うことで、多数の単語が取り扱い可能となる。

【0130】請求項5、12の発明では、連鎖確率が最小となる単語列を選択することで、聴覚心理実験で使用する無意味文を作成することができ、作成された文も恣意的な要素を含むことがない。

【0131】請求項6、13の発明では連鎖確率が最大となる単語列を音声認識候補の中の最適候補とするの  
10

で、従来の単語のみの音声認識に比べて、単語間の連鎖が考慮されており、音声認識精度が向上する。

【図面の簡単な説明】

【図1】本発明第1実施形態の言語処理装置の機能構成を示すブロック図である。

【図2】本発明第1実施形態の言語処理装置の処理手順を示すフローチャートである。  
20

【図3】本発明第1実施形態の言語処理装置の処理手順を示すフローチャートである。

【図4】本発明第2実施形態の言語処理装置の機能構成\*

\*を示すブロック図である。

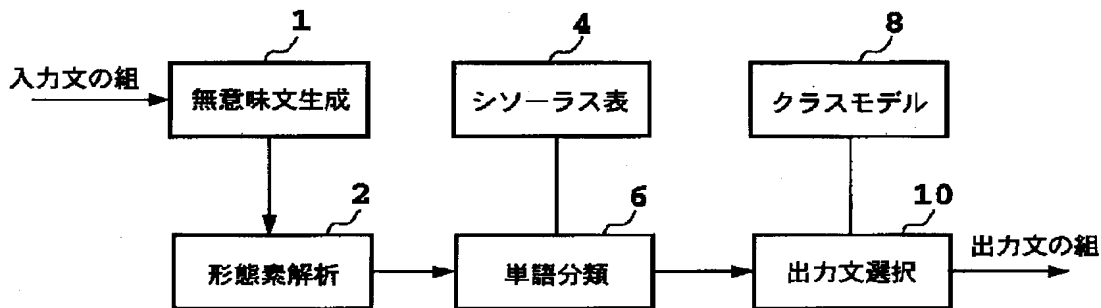
【図5】本発明第2実施形態の言語処理装置の処理手順を示すフローチャートである。

【図6】本発明第2実施形態の言語処理装置の処理手順を示すフローチャートである。

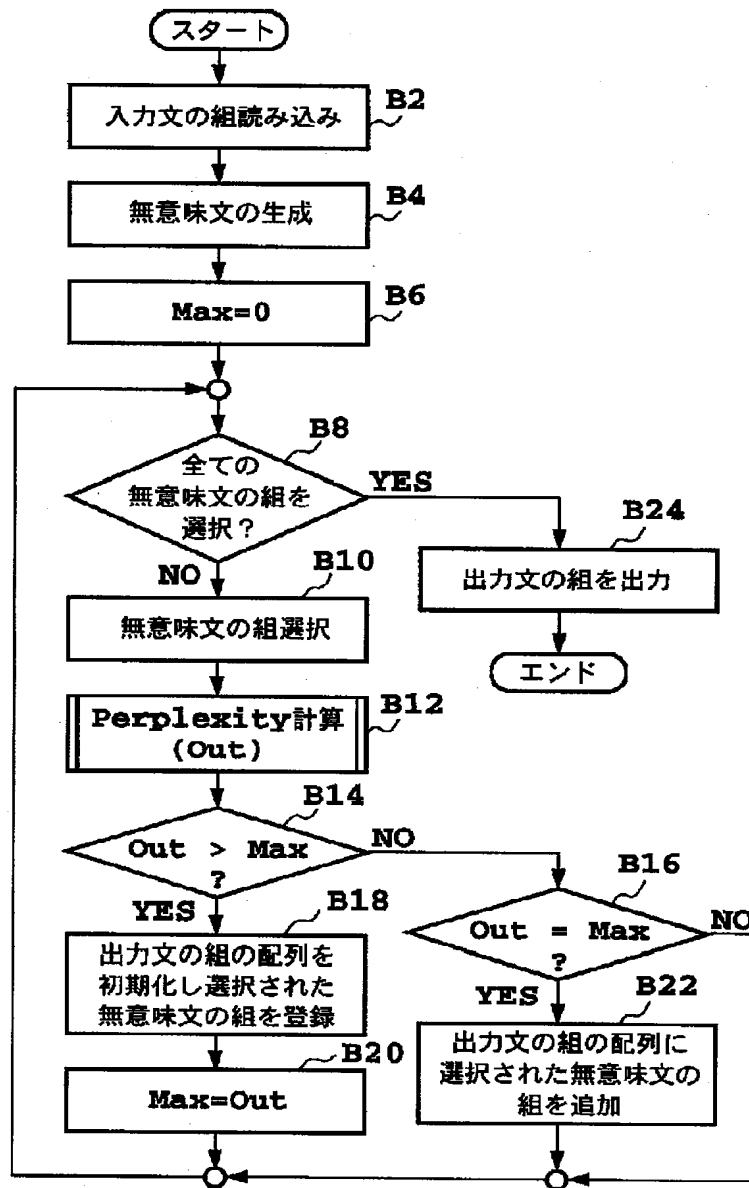
【符号の説明】

- 1 無意味文生成ブロック
- 2 形態素解析ブロック
- 4 シソーラス表格納ブロック
- 6 単語分類ブロック
- 8 クラス間のトライグラム確率の表を格納する処理ブロック
- 10 無意味文の組を選択する処理ブロック
- 22 形態素解析ブロック
- 24 単語トライグラム確率格納ブロック
- 26 単語連鎖確率を読み出す処理ブロック
- 28 シソーラス格納ブロック
- 30 単語分類ブロック
- 32 クラス間のトライグラム確率とクラス-単語確率の表を記憶する格納ブロック
- 34 クラストライグラムを計算する処理ブロック
- 36 クラストライグラム確率を利用して認識候補を選択する処理ブロック

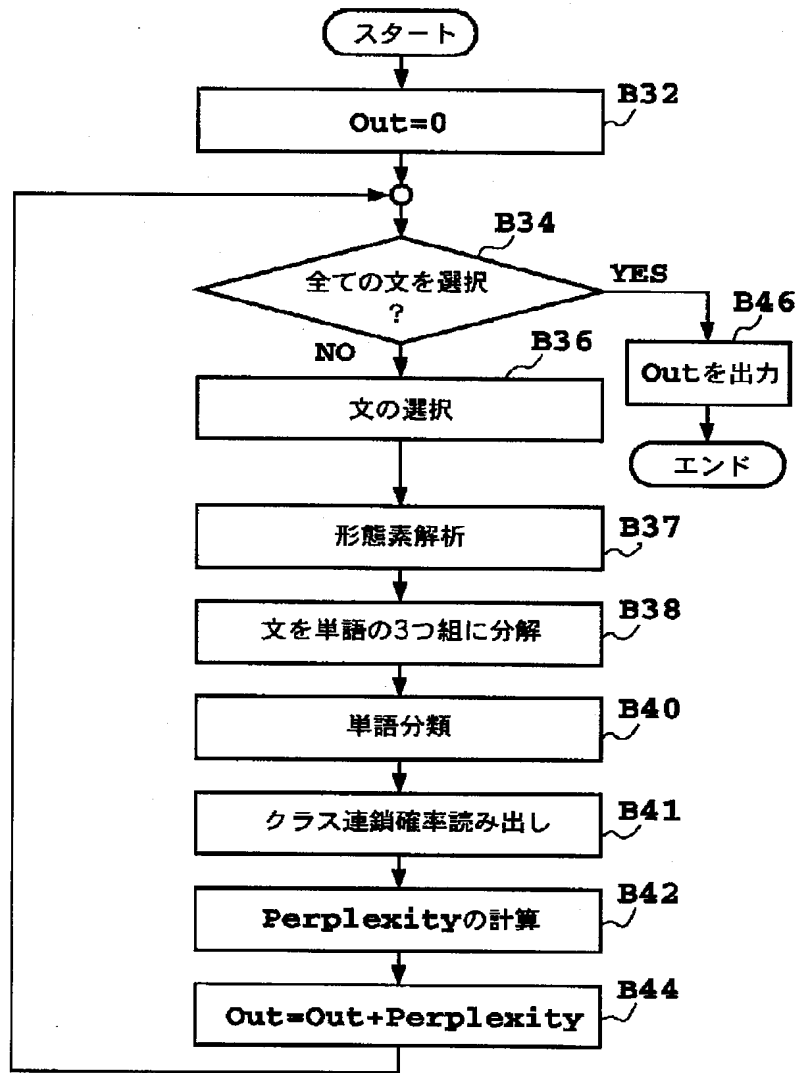
【図1】



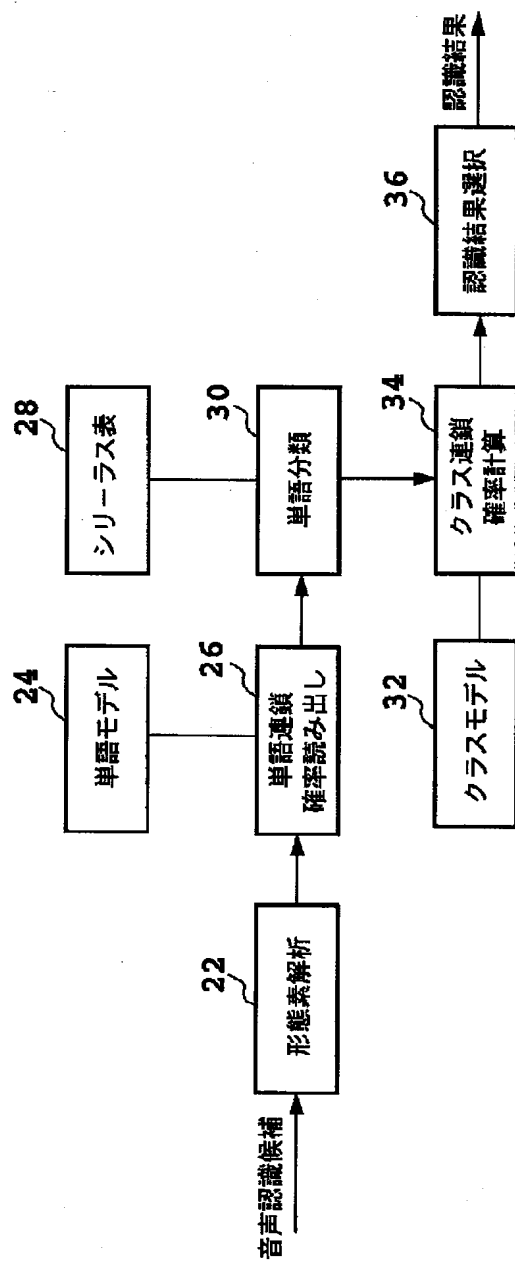
【図2】



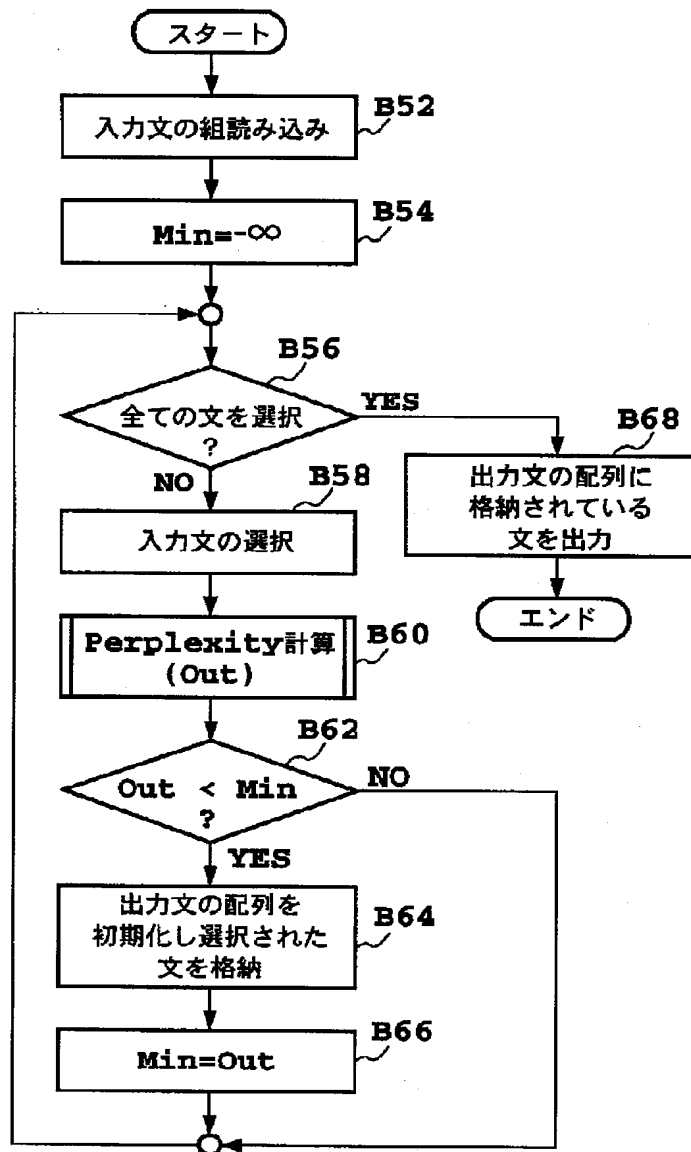
【図3】



【図4】



【図5】



【図6】

